Faculty Working Papers

MAD COST ESTIMATION

Soong H. Park. Assistant Professor, Department
of Accountancy
Rene P. Manes, Professor, Department of Accountancy

#709

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 5, 1980

MAD COST ESTIMATION

Soong H. Park. Assistant Professor, Department of Accountancy
Rene P. Manes, Professor, Department of Accountancy

#709

Summary

Economists and accountants have followed the lead of other scientists in resorting to ordinary least squares (OLS) regression for the estimation of costs. In so doing they have not given adequate consideration to the loss function associated with errors of OLS cost estimates. A different approach to cost estimation having a different loss function, i.e., the minimization of absolute deviation (MAD) is suggested. The technique for MAD estimation is shown to be a simplified version of an L.P. goal programming model. A brief discussion compares the estimation statistics of OLS and MAD approaches.

MAD Cost Estimation

Statistical cost estimation techniques have been introduced into the managerial accountant's tool kit for the last two decades. The main use of these tools has been in cost recognition. Management accountants derive estimated cost relationships in order to establish standard costs and to provide budget officers with predictive models. Either because the applications of statistical tools have been straightforward or because much of the data utilized is of a proprietary, in-house nature, relatively few of these studies have appeared in the accounting literature.[1] In those studies which have been published, the only technique applied and reported is that of ordinary least squares (OLS) regression analysis [e.g., McClenon, 1963; Benston, 1966; Comiskey, 1966].

Before we suggest an alternative to OLS regression analysis, it may be useful to recall a few facts about the origins of OLS. The method is attributed either to LeGendre or to Gauss, who contested with each other its first use back in the 18th century.[2] Gauss applied OLS to interpret astronomical data and defended the method because its linear estimator had the property of unbiasedness and minimized the variance of the estimators (Gauss-Markov Theorem). Throughout the 19th century, in genetics, biology, agriculture, etc. scientists resorted to OLS to infer the true nature of their sciences.

---

[1]In 1960 J. Johnston reported on empirical, econometric studies done for a number of industrial and financial sectors, Statistical Cost Analysis, McGraw Hill.

[2]Plackett, R. L., "The Discovery of the Method of Least Squares," Biometrika, 1972, 59, pp. 239-51.

The general approach of these men, as they relied on classical statistical methods in general and on OLS in particular, has been characterized as that of inference. The inference school, led by R. A. Fisher, considered statistics as a means of processing data into scientific relationships so that, by use of observations and experimentation, uncertainty about these relationships might be reduced in an unbiased manner.[3] Inference, thus simply dealt with information and did not allow itself to be influenced by the implications of the conclusions reached by the observer. For example, a statistician helping a biologist examine the bacterial density of a reservoir, could not, qua statistician, permit his concern about consequences of certain infestation levels to affect how his calculations would be made.

However in the last thirty years a new approach, that of decision theory, arose to challenge inference. Ferguson [1976] has summarized the four basic elements of this decision model:

1) The space Θ of states of nature, one of them "true" but unknown to the statistician.

2) The space A of action available to the decision maker.

3) The loss function $L(\theta,a)$ representing the loss for taking action $a \varepsilon A$ when the true state of nature is $\theta \varepsilon \Theta$.

4) An experiment yielding observation X, the distribution of which depends on the true state of nature, and which will help minimize $L(\Theta,a)$ from taking action a.

---

[3] Another major reason for scientific reliance on OLS is alleged to have been the ease of computation (no minus signs, differentiability, etc.). Whether this explanation is valid or not, the authors in their brief survey of the history of statistics have found very little evidence for this claim.

It is interesting to note that points 2) and 3) required specifications which the proponents of inference had been unwilling to make and that statisticians henceforth did build into their methodology a concern for consequences. Perhaps decision theory's greatest contribution has been to define a loss function explicitly, for without a loss function it has been impossible to suggest action to be taken under uncertainty such as to minimize detrimental consequences (or such as to maximize utility) among all possible states of nature.[4]

In fact the technique of OLS, a well established tool of inference, implies a loss function, and one which has been described as an epistemic loss function, i.e., a purely intellectual or cognitive one. We shall argue that economists, and subsequently accountants, in their increasing reliance on OLS for cost estimation, have not given adequate thought to the nature of this loss function nor given much thought to a decision theoretic approach to cost estimation, an activity which is engaged in primarily for decision making purposes.

This paper will examine the loss function implicit in the ordinary least squares (OLS) estimation techniques and offer an alternative technique of estimation based on a different loss function and on a more explicit statement of the costs of misestimation. This paper does not assert the superiority of one method over the other. It merely attempts to help arrive at a better understanding of the assumptions of regression analysis and to offer a method of computation when an alternative class of loss function seems more appropriate for a given situation.

---

[4]T. S. Ferguson, On the History of Statistics and Probability, edited by D. B. Owen, "Development of the Decision Model," pp. 335-6. The above summary relies heavily on Ferguson's chapter.

In order to facilitate the discussion we will use as an example a problem which was included in the December, 1974 CMA examination:

> The Ramon Co. manufactures a wide range of products at several different plant locations. The Franklin Plant, which manufactures electrical components, has been experiencing some difficulties with fluctuating monthly overhead costs. The fluctuations have made it difficult to estimate the level of overhead that will be incurred for any one month.
>
> Management wants to be able to estimate overhead costs accurately in order to plan its operation and financial needs better. A trade association publication to which Ramon Co. subscribes indicates that, for companies manufacturing electrical components, overhead tends to vary with direct-labor hours.
>
> One member of the accounting staff has proposed that the cost behavior pattern of the overhead costs be determined. Then overhead costs could be predicted from the budgeted direct-labor hours.
>
> Another member of the accounting staff suggested that a good starting place for determining the cost behavior pattern of overhead costs would be an analysis of historical data. The historical cost behavior pattern would provide a basis for estimating future overhead costs. The methods proposed for determining the cost behavior pattern included the high-low method, the scattergraph method, simple linear regression, and multiple regression. Of these methods Ramon Co. decided to employ the high-low method, the scattergraph method, and simple linear regression. Data on direct-labor hours and the respective overhead costs incurred were collected for the previous two years. The raw data follow:

	19_3		19_4	
	Direct-Labor Hours	Overhead Costs	Direct-Labor Hours	Overhead Costs
January	20,000	$84,000	21,000	$86,000
February	25,000	99,000	24,000	93,000
March	22,000	89,500	23,000	93,000
April	23,000	90,000	22,000	87,000
May	20,000	81,500	20,000	80,000
June	19,000	75,500	18,000	76,500
July	14,000	70,500	12,000	67,500
August	10,000	64,500	13,000	71,000
September	12,000	69,000	15,000	73,500
October	17,000	75,000	17,000	72,500
November	16,000	71,500	15,000	71,000
December	19,000	78,000	18,000	75,000

Using linear regression, the following data were obtained:

Coefficient of determination	.9109
Coefficient of correlation	.9544
Coefficients of regression equation	
Constant	39,859
Independent variable	2.1549
Standard error of the estimate	2,840
Standard error of the regression coefficient for the independent variable	.1437
True t-statistic for a 95% confidence interval (22 degrees of freedom)	2.074

The problem asks the students to construct the overhead cost function using the results of OLS computations and to defend the superiority of the OLS technique to the less sophisticated techniques such as HI-LO, visual curve fitting, etc. The cost pattern estimated in the problem by OLS is

$$\begin{aligned} y_i &= a + bx_i + e_i \\ &= \$39{,}859 + \$2.159x_i + e_i \end{aligned}$$

where y_i the dependent variable is the overhead cost of month i

x_i the independent variable is the direct labor hours for month i

e_i is the random error of the estimate for month i

a the intercept, is the estimated "fixed" overhead for one month

b the slope, is the estimated "variable" overhead for one hour of direct labor

In general economists, starting from the production function, posit a cost function $y = f(q)$ in which cost is a function of output q. Accountants more often define cost equations in which cost, $y = g(x_1, x_2, \ldots, x_n)$, where x represent inputs. The above problem is of the latter nature, i.e., a cost equation. In either case, because of the stochastic nature of the cost-volume relationships under examination and/or because of technical limitations in model building and measurement, economic or cost accounting forecasts (budgets) do not correspond exactly to cost observations. Statistical techniques will help in the assessment of the "true" underlying relations between inputs and outputs and their cost. Given the randomness of the cost relationship and measurement errors, no statistical estimates will yield error-free cost functions.

In a decision theory context, estimates $(\hat{y}_k)$ are predictions (signals) of uncertain states of nature, which will be used in evaluating alternative actions. Accordingly the estimated cost functions (or equations) $[y = f(q)]$ or $[y = g(x)]$ can be viewed as information systems. In general, perfectly accurate predictions will lead to better decisions than will erroneous ones. But, as indicated earlier, a perfect cost function would be unobtainable, even if it existed. Further, the consequences of various prediction errors are dependent upon the decision problem at hand and cannot be generalized. Since cost functions are estimated to facilitate decision making, the criterion for the parameter values of the cost function should be compatible with the decision objective, stated as a loss function.

We can formulate the cost function estimation problem as an optimization problem. The objective of the problem is to select parameter values such that the consequences of the differences between the estimated values and the observed values are at minimum. The manner in which the differences are measured, weighed and accumulated will determine the specific form of the objective function in the optimization problem.

Thus one can argue that the structure of the objective function in the estimation problem is directly related to the loss, or utility function in the decision problem. The following analysis is based on the assumption that the selection of the objective function can be guided by the knowledge of the loss function in the decision problem. We suggest that the loss function related to cost estimation be linear, (instead of quadratic) and this paper will therefore offer a family of objective functions which are linear.

Closer examination of the OLS method implies that the following loss function is being used in the determination of cost behavior:

(1) the penalties, $c(e_i)$, associated with positive or negative errors, are identical. That is, the magnitude, not the direction of error, is the sole determinant of the penalty:

$$c(e_i) = c(e_j) \text{ if and only if } |e_i| = |e_j|$$

(2) The relative penalties of the errors are the squares of the relative magnitude of the errors:

$$c(e_i)/c(e_j) = (|e_i|/|e_j|)^2$$

The above loss function means, among other things, that

(1) The same magnitude of unfavorable and favorable variances is equally significant in a standard costing system; or for a pricing decision, and

(2) A $500 variance is 25 times worse than a $100 variance due to the squaring aspect of the relative magnitude of the errors.

Of course, there are many situations in which neither economists nor accountants will be satisfied with the above implications. As a matter of fact, there is no reason to believe that all cost estimators would consciously choose such a quadratic loss function. The issue then is: how can we estimate costs by a technique having a loss function other than the quadratic one? The following section presents a computational technique with a set of alternative loss functions that can be used.

Cost Estimation by Minimization of Absolute Derivatives (MAD)

The criteria proposed for cost estimation as an alternative to OLS are the following:

(1)' The direction of errors, as well as the magnitude of errors, is pertinent in determining the cost of errors. Formally stated:

$$c(e_i) = kc(e_j)\ , \quad e_i = -e_j$$

where k is a positive weighting factor.

(2)' The relative penalty of the errors is proportional to the relative magnitude of the errors:

$$c(e_i)/c(e_j) = e_i/e_j \qquad e_i \times e_j \geq 0$$

A regression model which estimates coefficients of the cost equation and which at the same time satisfies the above criteria is based on a simplified version of the l.p. goal programming model. Assume the cost equation takes the traditional form.

$$y_i = a' + b'x_i + e_i' \qquad \text{for i observations} = 1, 2, \ldots, 24$$

The values of the coefficients, a' and b', must be estimated so as to minimize the sum of the absolute values of the error terms, e_i', viz [Wagner, 1959].

$$\text{Min} \sum_{i \varepsilon I} |a_j' + x_{ij}b_j' - y_i|$$

where x_i and y_i are constraints (tuples of observation values), a'_j and b'_j are the activity levels to be developed by the algorithm and the vertical strokes mean the absolute value of the expression they enclose.[5]

It is important to understand that a' and b', coefficients of the cost equation, have now become activity variables of the L.P. program; also that x, the independent variable, and y, the dependent variable of the cost equation have become constants in the objective function.

If deviations, e'_i, can be related to observation tuples, it is clear by definition that the objective function can be restated as

$$\text{Min} \sum_{i \varepsilon I} [e_i^+ + e_i^-] \qquad \text{for } e'_i = e_i^+ + e_i^-.$$

And indeed, Charnes, Cooper and Ferguson[6] have shown the equivalence in the linear programming format of the above objective function developed by Wagner and of

$$\text{Min} \sum_{i \varepsilon I} [e_i^+ + e_i^-]$$

$$\text{s.t. } a'_j + x_{ij} b'_j - e_i^+ + e_i^- = y_i \text{ for } i = 1, 2, ---, n$$

$$a', b', e_i^+, e_i^- \geq 0$$

---

[5]Concerned by overemphasis on outliers, Sharp [1971] applied the same MAD criterion in security risk analysis. For highly diversified portfolios, he found differences in OLS & MAD to be relatively small and reached the tentative conclusion that gains from use of MAD would be modest.

[6]Charnes, A., Cooper, W. W. & Ferguson, R. O., "Optimal Estimation of Executive Compensation by Linear Programming," Management Science 1 (1955) pp. 138-51.

where e_i^+ represents all overage or deviation above the regression line, and e_i^- all underage or a deviation beneath the regression line. This formulation will be recognized as the goal programming first developed by Charnes & Cooper,[7] introduced to the accounting literature by Ijiri[8] and made the subject of book length study by Sang M. Lee [1972].

This model is a very simplified version of goal programming because there is no problem with goal dimensionality (all deviations are measured in \$) and because, therefore, no preemptive priorities are established for the goals (i.e., the observations to be fitted to a regression equation) or applied to the deviations from these goals.

Referring to the sample problem, the constraint set can be further identified as follows

$$a_1 - a_2 + x_i b_1 - x_i b_2 - e_i^+ + e_i^- = y_i$$

or

(1) $$a_1 - a_2 + 20{,}000 b_1 - 20{,}000 b_2 - e_1^+ + e_1^- = 84{,}000$$

(2) $$a_1 - a_2 + 25{,}000 b_1 - 25{,}000 b_2 - e_2^+ + e_2^- = 99{,}000$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

(24) $$a_1 - a_2 + 18{,}000 b_1 - 18{,}000 b_2 - e_{24}^+ + e_{24}^- = 75{,}000$$

---

[7] Charnes, A. & Cooper, W. W. Management Models and Industrial Applications of Linear Programming (Wiley 1961) refer also "Goal Programming & Multiple Objective Optimization," European Jour. of O. R. (1977) pp. 39-54.

[8] Ijiri, Y. Management Goals and Accounting for Control (Amsterdam: North Holland 1965), reference also L. N. Killough and T. L. Souders, "Goal Programming for Public Accounting Firms," The Accounting Review April 73, pp. 268-279.

The cost equation coefficients, a' and b', which are now the decision variables of the l.p. model, have been replaced by pairs (a_1,a_2) and (b_1,b_2) because of the linear programming requirement that the values of all variables be non-negative. Only one variable in each pair will take on a value while the other remains at zero. If the intercept is positive a_1 will assume that value whereas a_2 will take on the intercept value if it is negative. Likewise b_1 and b_2 represent positive and negative slopes respectively. Finally the error terms e_i^+ and e_i^- represent positive or negative deviation terms associated with the i^{th} set of observations. Again only one of these terms per equation can take on a value. Stated in more familiar notation

$$\text{Minimize } x_{29} + x_{30} \qquad \text{(Prob. 1)}$$

$$\text{s.t.} \begin{cases} (1) \quad x_1 - x_2 + 20{,}000x_3 - 20{,}000x_4 - x_5^+ + x_5^- = 84{,}000 \\ (2) \quad x_1 - x_2 + 25{,}000x_3 - 25{,}000x_4 - x_6^+ + x_6^- = 99{,}000 \\ \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ (24) \; x_1 - x_2 + 18{,}000x_3 - 18{,}000x_4 - x_{28}^+ + x_{28}^- = 75{,}000 \end{cases}$$

and

$$\text{s.t.} \qquad - x_5^+ - x_6^+ - \ldots - x_{28}^+ + x_{29} = 0$$

$$\text{and} \qquad - x_5^- - x_6^- - \ldots - x_{28}^- + x_{30} = 0$$

$$\text{all } x_i\text{s} \geq 0$$

The first 24 constraints have the form [D]. $\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ e_i^+ \\ e_i^- \end{bmatrix} = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{bmatrix}$

$$\text{where } [D] = \begin{bmatrix} 1 & -1 & +\text{obs.}1 & -\text{obs.}1 & -1 & +1 & & & & \\ 1 & -1 & +\text{obs.}2 & -\text{obs.}2 & & & -1 & +1 & & \\ \cdot & & \cdot & & & & & \cdot & & \\ \cdot & & \cdot & & & & & & \cdot & \\ \cdot & & \cdot & & & & & & & \cdot \\ 1 & -1 & +\text{obs.}24 & -\text{obs.}24 & & & & & -1 & +1 \end{bmatrix}$$

The last two constraints aggregate the positive and negative errors respectively. While not necessary, this formulation will provide a convenient way of weighting positive and negative errors differentially. If we choose to, we can multiply x_{29} and x_{30} by different weights.

For our example as shown in Table 1 and Figure 2 the values of a' and b' are +40,170 and +2.167 respectively, which are quite similar to that of OLS estimates. In the next section we shall show that for a symmetric distribution of error terms, this result should be expected.

In the usual goal programming problem the several goals of the optimizer which become part of the constraint set and which may conflict with one another represent operational conditions [refer S. M. Lee, Chapters 8 to 14, which refer to production planning, marketing, corporate planning, medical care planning, etc.]. The separate constraints (goals) may or may not have common dimensions, and when they do not, they can only be ranked ordinally, i.e., by preemptive priorities. Such priorities are then assigned to the slack in the goal constraints (deviation variables) so that deviations can be summarized in a prescribed order. Within the particular subsets of constraints classified under one priority, additional refinement of goal attainment can be achieved by the assignment of different weights to deviated variables.

In our cost estimation goal program, the goals are all cost fitting statements of the same dimension and typically one set of observations is no more important than the next in contributing to the definition of the cost relationship. For this reason (and because of the single monetary dimension), no preemptive priorities need to be established and the cost estimation process will be deliberately indifferent to fitting one constraint instead of another.[9] On the other hand, negative deviations may be viewed more seriously than positive ones (or vice versa) and weighted accordingly with a corresponding effect on the MAD loss function. The objective function would then become

$$\text{Min } wx_{29} + vx_{30} \qquad \text{for } w \text{ \& } v\text{, subjective weightings of the positive and negative deviations.}$$

For example, suppose a company is considering a bid on a job. If cost estimation is too low (positive error), the company suffers a reduction in profit. If the estimation is too high (negative error), the company unnecessarily reduces its chances of obtaining the job.

---

[9] Ignizio [1978] provides a nuclear engineering cost example, in which certain parameter values must fall within a given range due to physical limitations. He thus assigns preemptive priorities to the objective of estimating these parameters over others.

Let us assume that management of the firm assesses that the penalty or cost associated with positive errors (underestimation) is twice as severe as that associated with negative errors (overestimation). In such a case, the objective function for estimation purposes can be adjusted as follows:

$$\text{Minimize} \quad x_{29} + 2x_{30} \qquad \text{(Prob. 2)}$$

Constraints are same as Problem 1.

The estimated coefficient values for this problem are an intercept (a') of \$42,750 and a slope coefficient (b') of \$2.062 per unit (see Table 1 and Figure 1).

On the other hand if management wishes to use past data for the purpose of setting standards, it may for motivational reasons wish to set up a tight but still attainable standard. In this circumstance, it is more plausible to consider the following objective function:

$$\text{Minimize} \quad kx_{29} + x_{30}, \qquad k > 1 \qquad \text{(Prob. 3)}$$

Constraints are same as in Problem 1.

For $k = 2$ we obtain the estimated intercept (a') of \$38,170, and coefficient (b') unit cost of \$2.167 (see Table 1).

Such a L. P. goal programming formulation can easily be adjusted to more than one independent variable (multiple regression). It can also accomodate indicator variables to account for effects not reflected in the values of the independent variables, as well as allowing for piece-wise-linear equations (see table 1). We can force the fitted line (plane) to pass through a given desired point by any one of several means. For example, if we believe that any given tuple of

(x_i, y_i), exactly represents the cost relationship and we wish the regression to pass thru that point[10], we can exclude the deviation for that particular constraint from the aggregation of positive and negative summation, include them separately in the objective function appropriately weighted by M;

$$\text{Min } wx'_{29} + vx(30)' + Mx_i^+ + Mx_i^-$$

where x'_{29} and x'_{30} are the aggregation of positive and negative deviations excluding the ith observation set.

In the above case, due to the weighting factor, the estimates of the MAD estimates will not be the same as OLS estimates even if the data were distributed such that the errors terms satisfy the OLS assumptions, i.e., of independent and identical normal distributions.

Non-Symmetric Error Distribution

Next, we consider the case of non-normal, non-symmetric error distributions. The measure of central tendency that minimizes the sum of the squared deviations is the mean or the average (OLS estimates) [Gauss, 1821], whereas the measure that minimizes the sum of absolute deviations is the median of the distribution (Minimum Absolute Deviation (MAD) estimates) [Laplace, 1812]. For symmetric distributions, the mean and the median are the same; estimates of parameters will tend to be similar where the distribution of error terms is symmetric and then we would expect the OLS and MAD estimates

---

[10] If that point is given by the set of sample means, we can add to the constraint set the linear restriction $a + b_1\bar{x}_1 \ldots b_m\bar{x}_m = \bar{y}$ (Wagner, 1959, p. 208).

to be quite similar.[11] The roles that these central tendency measures play have been discussed in the accounting literature by Barefield [1969] and others [Peterson and Miller, 1964]. Our discussion is not to repeat the properties of the measures, but to demonstrate how those measures can be estimated using the MAD estimation technique. For convenience, we will use a non-symmetric triangular distribution of error terms as the example for analysis. The range of distribution is (0,3) and the peak point is 2 as shown in figure 3. The mode of the distribution is 2 and the mean and median are 5/3 and 1.732 respectively (Details are shown in Appendix).

<u>Insert Figures 3 and 4 about here</u>

Given this distribution, the estimator which minimizes the sum of squared errors (Σe_i^2) is the mean, 5/3; the sum of absolute deviations ($\Sigma|e_i|$) is the median, 1.732; and the mode can be estimated using the weighted linear loss function, $\Sigma|e_i| + w|e_i|$ (again see Appendix for details).

The point to be stressed is that the MAD procedure can be used either to reflect the non-quadratic loss functions (figure 1) or to estimate various central measures when the errors are not distributed symmetrically (figures 2 and 3). Table 2 summarizes these alternatives.

<u>Insert Table 2 about here</u>

---

[11] However, Wilson [1978] found that when outliers are present MAD estimates are significantly more efficient than OLS estimates.

We have demonstrated how the values of the coefficients can be obtained. The key remaining question is how good or useful are the estimates? This issue is addressed in the next section.

Evaluation of the Results

One obvious measure of the "goodness" of estimation is the question of how much of the variations in overhead costs is explained by the estimated cost pattern? To answer this question for MAD we shall start by defining the term total variation.

The OLS estimates minimizes the sum of the squared errors. Therefore, the total variations should also be expressed in terms of squared deviations. Had the direct labor hours information not been available to us, the estimate of the overhead cost that minimizes the sum of squared errors would be the mean of the overhead costs: $\Sigma(y_i - y_{mean})^2$. This measure of variation is called the Total Sum of Squares (TSS).

However, since the objective of MAD estimation is to minimize the total absolute deviation (TAD), a reasonable measure of variation would be the sum of deviations from the median:

$$TAD = \Sigma|y_i - y_{median}|.$$

We can obtain the corresponding measures of the variations that have been, or have not been, accounted for by introducing direct labor hours as an explanatory variable. For OLS estimation it is the Sum of Squared Errors ($SSE = \Sigma\, e_i^2$), and for MAD estimation it is the Sum of Absolute Deviations ($SAD = \Sigma|e_i|$), which is the value of the objective function in the L.P. formulation. We can then use these measures to obtain an index of the deviations explained by the estimated cost pattern.

Using Problem 1 as an example, for OLS regression:

$$R^2 = 1- \frac{SSE}{TSS} = 1 - \frac{\Sigma\, e_i^{\;2}}{\Sigma(y_i - y_{mean})^2} = .9101$$

and for MAD estimation:

$$Z = 1 - \frac{SAD}{TAD} = 1 - \frac{\Sigma|e_i|}{\Sigma|y_i - y_{med}|} = .6952.$$

Another measure that can be used for evaluation of the cost equation is the average error associated with the estimated costs. In OLS the average error is labeled the "standard error of the estimates (SE)."

$$SE = \frac{SSE}{d.f.} = \frac{\Sigma\, e_i^{\;2}}{d.f.} = \frac{1.77463E8}{22} = 2840.16$$

d.f. is degrees of freedom and

$1.77463E8 = 1.77463 \times 10^8$.

As discussed in the Appendix I, of the 24 observations due to the estimation technique only 22 error terms are free to take on values. Thus the sum of squared errors are divided by 22 to obtain the average. Similarly, the mean deviation associated with MAD estimates (which also contains 22 degrees of freedom) can be calculated as follows

$$MD = \frac{SAD}{d.f.} = \frac{\Sigma|e_i|}{d.f.} = \frac{55,170}{22} = 2507.7$$

One note of caution is in order. While it can be argued that the measures R^2, SE and Z, MD possess similar properties, these measures should not be used for the purpose of comparing the two estimation techniques. Both methods of estimation yield optimal solutions with respect to their given criteria. Choosing between the two methods should be based on the manager's judgment as to which set of criteria (OLS or MAD) is more appropriate for the situation at hand. That is, the choice of the estimation procedure (OLS or MAD) should be based on the basic assumptions of the estimation procedures, not based on the numbers resulting from a given problem; the latter are not suitable for comparison.

As noted in the introduction, the purpose of this paper has been to point out the inherent assumptions of OLS regression analysis, and to offer an alternative method of cost estimation should the MAD assumptions correspond more closely to needs of management than do the OLS assumptions. MAD estimation can be applied to more complex situations (e.g., multiple regression, indicator variables, etc.), but it should be recognized that much of the apparatus developed for OLS technique is not yet available for MAD. For example, the statistical properties of the estimates and the analysis of the error terms and the coefficients have not been developed as well as they have for OLS regression analysis. Should the actual application of MAD in real situations prove useful, we can expect future studies to develop the necessary understanding and techniques for further analysis.

As early as 1821, Gauss conceded that the choice of a loss function was somewhat arbitrary, and that Laplace's choice of absolute error was no more arbitrary than his own choice of squared error. However, when we cast the cost estimation problem in a decision setting, we have the basis for choosing an appropriate loss function. In this paper, we have shown how a linear programming model can be used for estimation of the cost patterns when the loss function is linear.

Appendix

This appendix is prepared to provide the computational details for an example used in this paper. The triangular distribution has the following density function:

$$f(x) = \begin{cases} hx & , \ a < x \leq c \\ h'(b - x) & , \ c < x \leq b \end{cases}$$

In the example a = 0, b = 3, and c = 2 which is mode. Then, the height of the triangle k is 2/3 and the slopes h and h' are 1/3 and 2/3 respectively. Initially we wish to find the value of the central tendency measures:

Mean: $E(x) = \int xf(x)dx$

$$= \int_a^c x[hx]dx + \int_c^b x[h'(b - x)]dx$$

$$= \int_a^c hx^2dx + \int_c^b [h'bx - h'x^2]dx$$

$$= 1/3(x^3|_a^c) + 1/2(h'bx^2|_c^b) - 1/3(h'x^3|_c^b)$$

$$= 5/3$$

Median: $p(x \geq m) = .5$

$$\int_a^m f(x)d_x = \int_a^m hxdx = .5$$

$$1/2(hx^2|_a^m) = 1/2hm^2 = .5$$

$h = 1/3$ thus

$$m^2 = 3 \quad m = 3^{1/2} = 1.732$$

The estimate that minimizes the sum of weighted absolute deviations.

Minimize $L = \int_a^b f(x)g(x)dx$ where $f(x)$ is the probability density function
$g(x)$ is the loss function

That is, find the estimate $\bar{e}$ such that the panalty

$$L = \int_a^{\bar{e}} f(x)g(x)dx + \int_{\bar{e}}^b f(x)g(x)dx \text{ is minimized.}$$

$$L = \int_a^{\bar{e}} hx(\bar{e} - x)dx + w \int_{\bar{e}}^c hx(x-\bar{e})dx + w \int_c^b h'(b - x)\ (x - \bar{e})dx$$

After solving this equation, we shall take partial derivatives to obtain the value of $\bar{e}$:

$$\frac{\partial L}{\partial \bar{e}} = 1/2[(h + wh)\bar{e}^2 - ha^2 - whc^2 - wh'b^2 + 2wh'bm - wh'm^2] = 0$$

for $a = 0$, $b = 3$, $c = 2$, $h = 1/3$, $h' = 2/3$ and $w = 1$,
$\bar{e}^2 = 3$ thus the estimator is the median.

In order to calculate the weight w which would yield the Mode as its estimator, we simply let, $a = 0$, $b = 3$, $c = 2$, $h = 1/3$, $h' = 2/3$ and $e = c$. Then $w = 2$ which implies that the loss function $g(x) = e_i + 2ej$ will yield the maximum likelihood estimator.

M/B/162

TABLE 1

Summary of the Estimated Cost Patterns

Technique	Criterion	Cost Equation	Total Errors	d.f.	Mean Error	Variation Explained
O.L.S.	Min Σe_i^2	$y_i = 39{,}859 + 2.1549x_i$	$\Sigma e_i^2 = 1.77463E8$	22	$\sqrt{\frac{\Sigma e_i^2}{d.f.}} = 2{,}840.16$	$1 - \frac{\Sigma e_i^2}{\Sigma(y_i - y_{mean})^2} = .9109$
$\Sigma(y_i - y_{mean})^2 = 1.9913E9$	Min Σe_i^2	$y_i = 56{,}906 + .9436x_i + 2.1028x_i'$	$\Sigma e_i^2 = 5.4828E7$	21	1,615.82	.9724
MAD	Min $\Sigma\|e_i\|$	$y_i = 40{,}170 + 2.167x_i$	$\Sigma\|e_i\| = 55{,}170$	22	$\frac{\Sigma\|e_i\|}{d.f.} = 2507.7$	$1 - \frac{\Sigma\|e_i\|}{\Sigma\|y_i - y_{med}\|} = .6952$
$\Sigma\|y_i - y_{med}\| = 181{,}000$	Min $\Sigma\|e_i\|$	$y_i = 59{,}750 + .75x_i + 2.25x_i'$	$\Sigma\|e_i\| = 30{,}250$	21	1440.5	.8329
Weighted MAD	Min $\Sigma e_{1,i} + 2e_{2,i}$	$y_i = 42{,}750 + 2.062x_i$	$\Sigma e_{1,i} = 42{,}870$ $\Sigma e_{2,i} = 13{,}687$			
		$y_i = 58{,}710 + .8571x_j + 2.357x_j'$	$\Sigma e_{1,i} = 26{,}430$ $\Sigma e_{2,i} = 6{,}429$			
	Min $\Sigma 2e_{1,i} + e_{2,i}$	$y_i = 38{,}170 + 2.167x_i$	$\Sigma e_{1,i} = 13{,}333$ $\Sigma e_{2,i} = 48{,}830$			
		$y_i = 55{,}500 + 1x_i + 1.917x_i'$	$\Sigma e_{1,i} = 5{,}917$ $\Sigma e_{2,i} = 25{,}580$			

Legends:

y_i: Total overhead cost for month i

x_i: Direct Labor Hours for month i

x_i': Excess Direct Labor Hours for month i

$$x_i' = \begin{cases} x_i - 17{,}000, & \text{for } x_i > 17{,}000 \\ 0, & \text{for } x_i \le 17{,}000 \end{cases}$$

$e_{1,i}$: the magnitude of over estimation for month i, i.e., $y_i + e_{1,i} = a + bx_i$

$e_{2,i}$: the magnitude of under estimation for month i, i.e., $y_i = a + bx_i + e_{2,i}$

The estimate that minimizes the sum of weighted absolute deviations.

Minimize $L = \int_a^b f(x)g(x)dx$ where $f(x)$ is the probability density function
$g(x)$ is the loss function

That is, find the estimate $\bar{e}$ such that the panalty

$$L = \int_a^{\bar{e}} f(x)g(x)dx + \int_{\bar{e}}^b f(x)g(x)dx \text{ is minimized.}$$

$$L = \int_a^{\bar{e}} hx(\bar{e} - x)dx + w \int_{\bar{e}}^{c} hx(x-\bar{e})dx + w \int_c^b h'(b - x)\ (x - \bar{e})dx$$

After solving this equation, we shall take partial derivatives to obtain the value of $\bar{e}$:

$$\frac{\partial L}{\partial \bar{e}} = 1/2[(h + wh)\bar{e}^2 - ha^2 - whc^2 - wh'b^2 + 2wh'bm - wh'm^2] = 0$$

for $a = 0$, $b = 3$, $c = 2$, $h = 1/3$, $h' = 2/3$ and $w = 1$,
$\bar{e}^2 = 3$ thus the estimator is the median.

In order to calculate the weight w which would yield the Mode as its estimator, we simply let, $a = 0$, $b = 3$, $c = 2$, $h = 1/3$, $h' = 2/3$ and $e = c$. Then $w = 2$ which implies that the loss function $g(x) = e_i + 2ej$ will yield the maximum likelihood estimator.

TABLE 1

Summary of the Estimated Cost Patterns

Technique	Criterion	Cost Equation	Total Errors	d.f.	Mean Error	Variation Explained
O.L.S.	Min Σe_i^2	$y_i = 39{,}859 + 2.1549x_i$	$\Sigma e_i^2 = 1.77463E8$	22	$\sqrt{\frac{\Sigma e_i^2}{d.f.}} = 2{,}840.16$	$1 - \frac{\Sigma e_i^2}{\Sigma(y_i - y_{mean})^2} = .9109$
$\Sigma(y_i - y_{mean})^2$ $= 1.9913E9$	Min Σe_i^2	$y_i = 56{,}906 + .9436x_i + 2.1028x_i'$	$\Sigma e_i^2 = 5.4828E7$	21	1,615.82	.9724
MAD	Min $\Sigma\lvert e_i\rvert$	$y_i = 40{,}170 + 2.167x_i$	$\Sigma\lvert e_i\rvert = 55{,}170$	22	$\frac{\Sigma\lvert e_i\rvert}{d.f.} = 2507.7$	$1 - \frac{\Sigma\lvert e_i\rvert}{\Sigma\lvert y_i - y_{med}\rvert} = .6952$
$\Sigma\lvert y_i - y_{med}\rvert$ $= 181{,}000$	Min $\Sigma\lvert e_i\rvert$	$y_i = 59{,}750 + .75x_i + 2.25x_i'$	$\Sigma\lvert e_i\rvert = 30{,}250$	21	1440.5	.8329
Weighted MAD	Min $\Sigma e_{1,i} + 2e_{2,i}$	$y_i = 42{,}750 + 2.062x_i$	$\Sigma e_{1,i} = 42{,}870$ $\Sigma e_{2,i} = 13{,}687$			
		$y_i = 58{,}710 + .8571x_i + 2.357x_i'$	$\Sigma e_{1,i} = 26{,}430$ $\Sigma e_{2,i} = 6{,}429$			
	Min $\Sigma 2e_{1,i} + e_{2,i}$	$y_i = 38{,}170 + 2.167x_i$	$\Sigma e_{1,i} = 13{,}333$ $\Sigma e_{2,i} = 48{,}830$			
		$y_i = 55{,}500 + 1x_i + 1.917x_i'$	$\Sigma e_{1,i} = 5{,}917$ $\Sigma e_{2,i} = 25{,}580$			

Legends: y_i: Total overhead cost for month i
x_i: Direct Labor Hours for month i
x_i': Excess Direct Labor Hours for month i

$$x_i' = \begin{cases} x_i - 17{,}000, & \text{for } x_i > 17{,}000 \\ 0, & \text{for } x_i \leq 17{,}000 \end{cases}$$

$e_{1,i}$: the magnitude of over estimation for month i
i.e., $y_i + e_{1,i} = a + bx_i$

$e_{2,i}$: the magnitude of under estimation for month i
i.e., $y_i = a + bx_i + e_{2,i}$

Table 2: Estimation Methods Suitable for Different Conditions

Estimation Method	Loss Function	Central Tendency*
OLS	e_i^2	Mean
MAD	$\lvert e_i \rvert$	Median
Weighted MAD	$w\lvert e_i \rvert + \lvert e_j \rvert$	Mode (MLE)

*Assuming non-symmetric independent and identical error distribution.

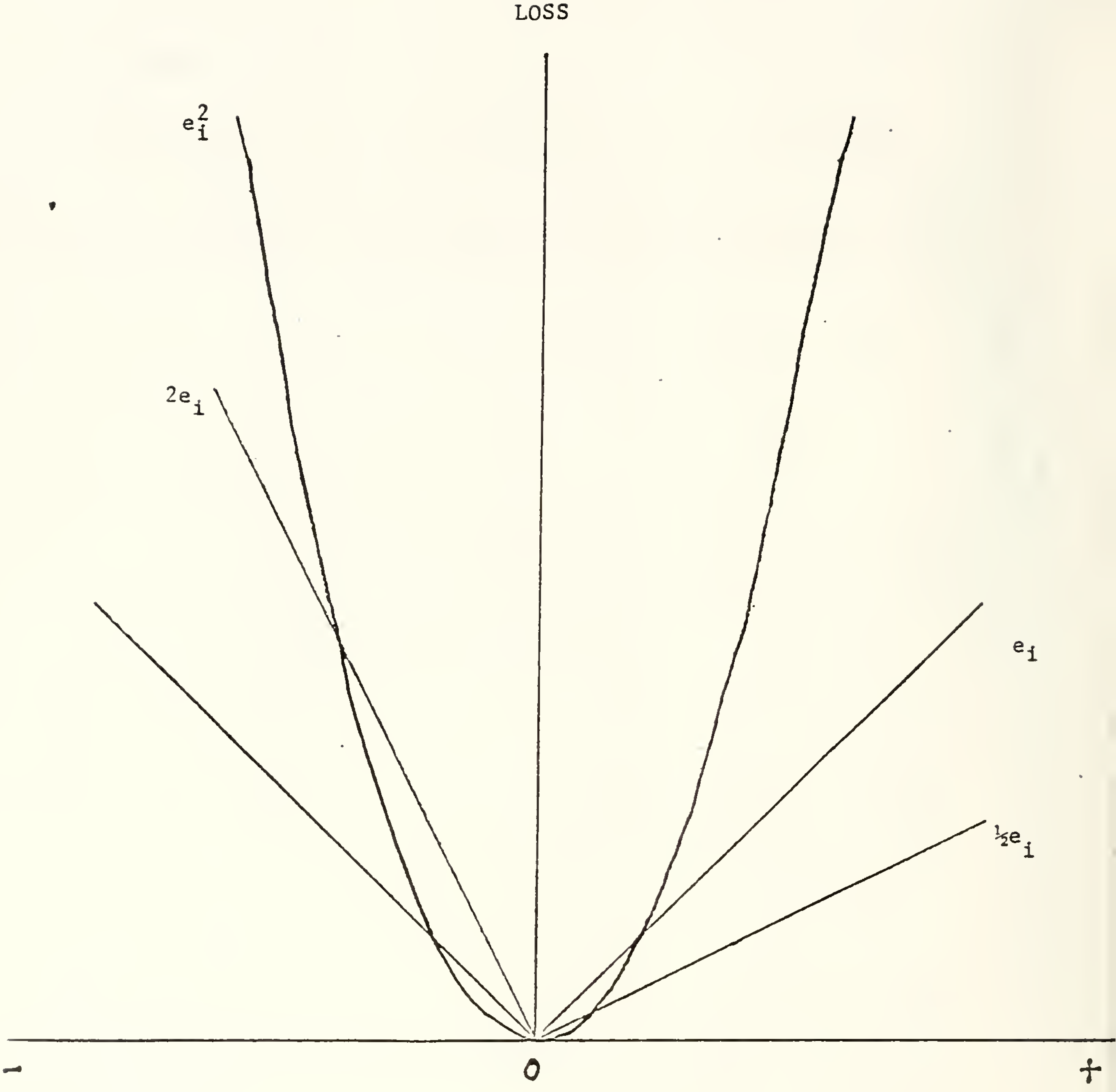


Figure 1

Alternative Loss Functions

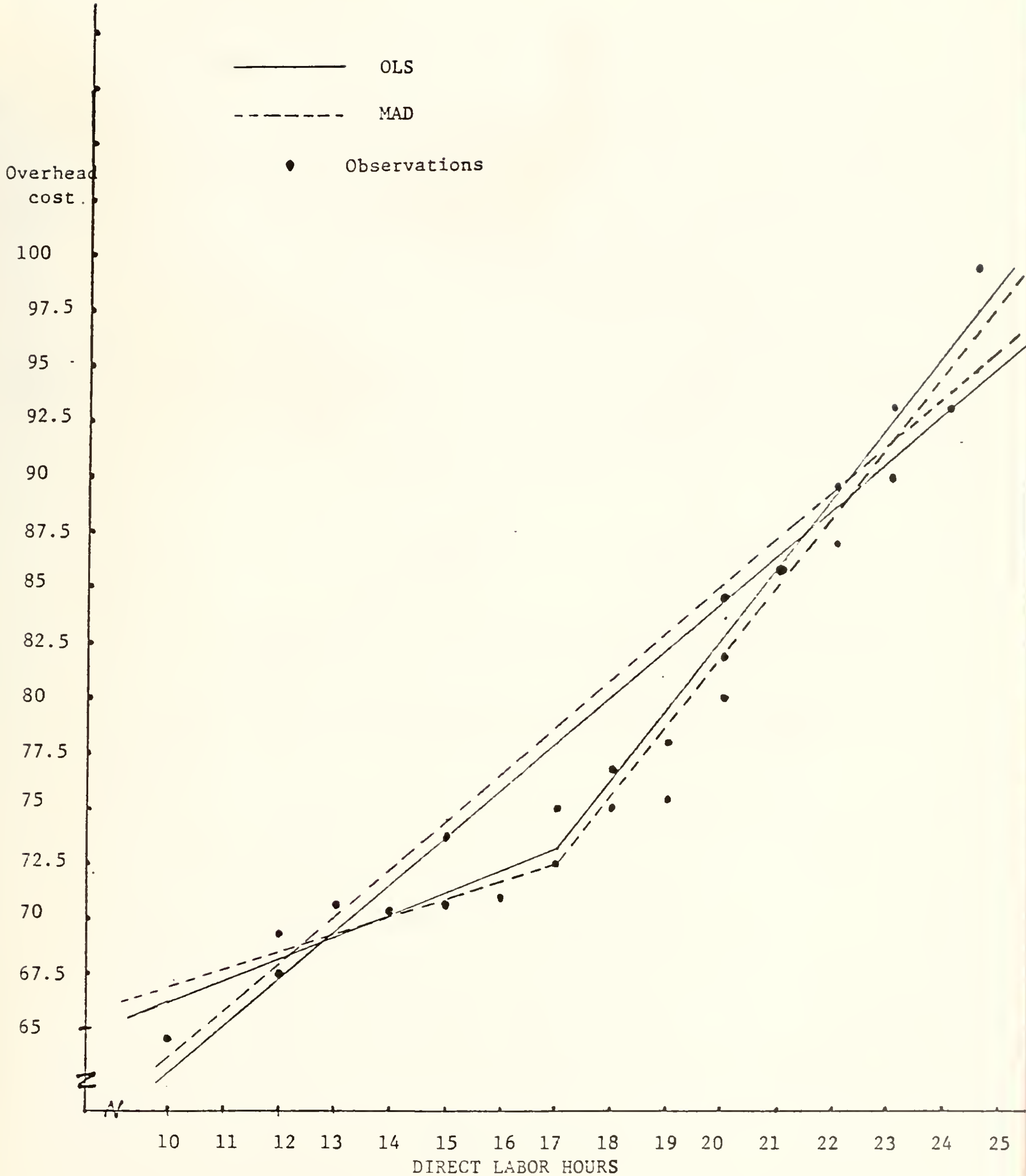
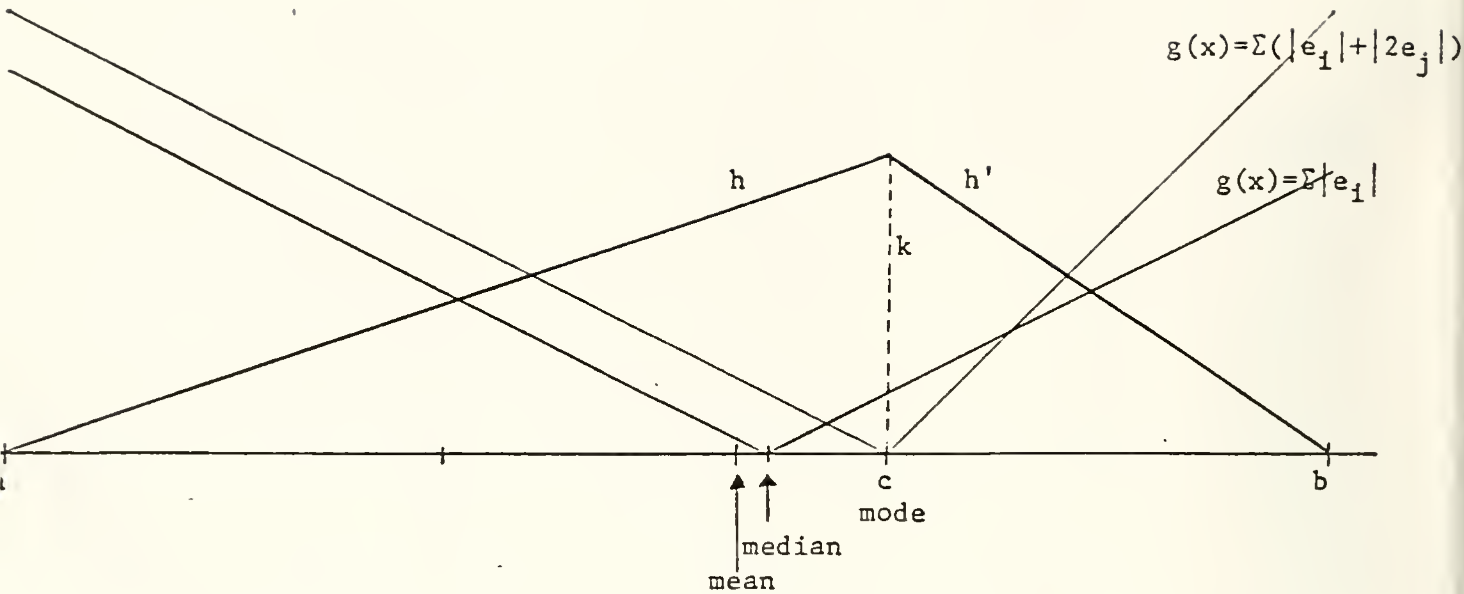


Figure 2

Estimated Cost Functions



$$f(x) = \begin{cases} hx & , a<x<c \\ h'(b-x) & , c<x<b \end{cases}$$

f(x): probability density function, g(x): loss function

Figure 3: A Non-Symmetrical Triangular Distribution

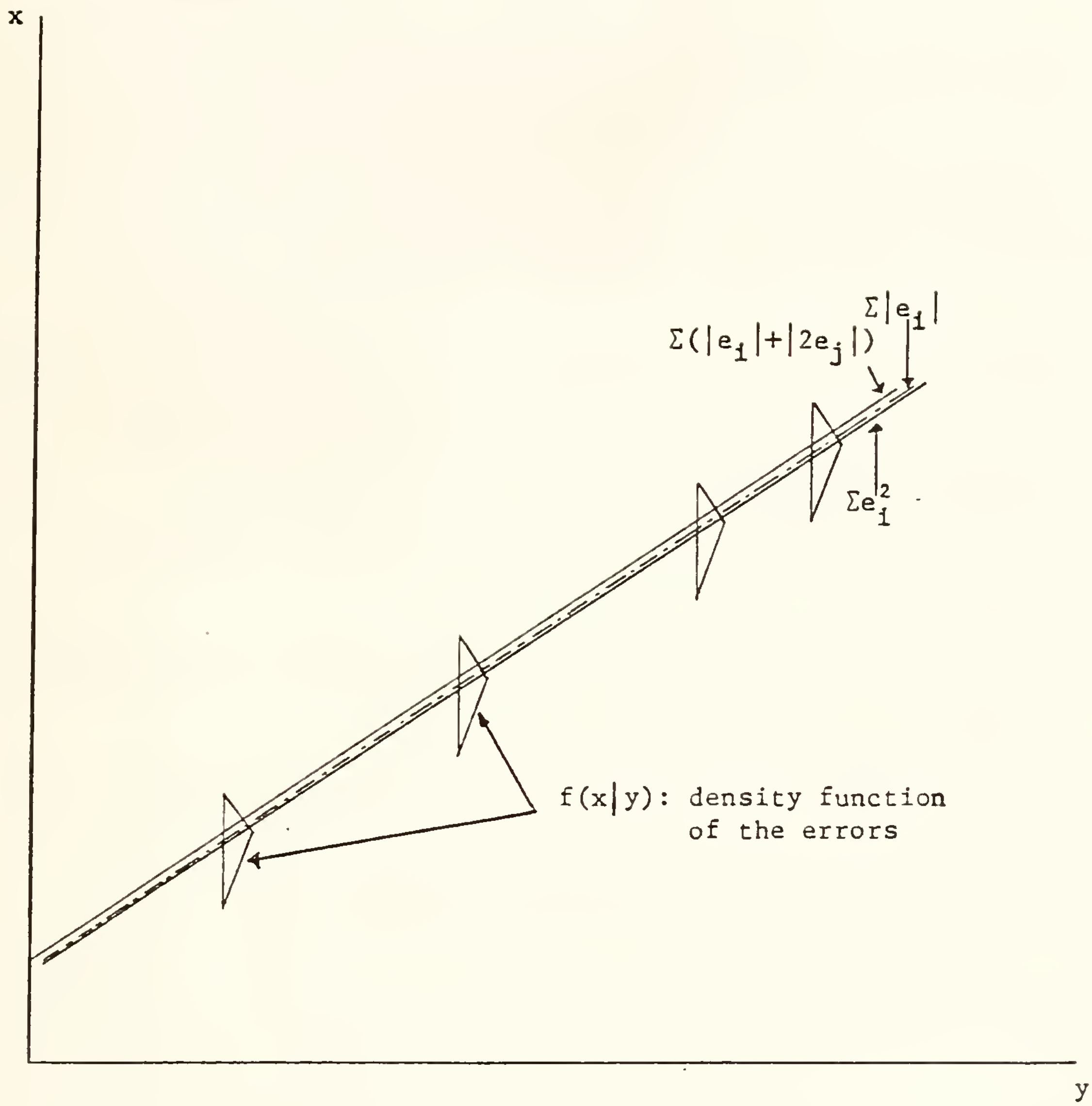


Figure 4: Cost Equations Under Various Loss Functions

REFERENCES

Barefield, R. M., "Comments on a Measure of Forecasting Performances," Journal of Accounting Research, Autumn, 1969, pp. 324-27.

Benston, G. J., "Multiple Regression Analysis of Cost Behavior," Accounting Review, October 1966, pp. 657-72.

Charnes, A. and Cooper, W. W., Management Models and Industrial Applications of Linear Programming, Wiley, 1961

______________, ______________, "Goal Programming and Multiple Objective Optimization," European J. of O. R., 1977, pp. 39-54.

______________, ______________, and Ferguson, R. O., "Optimal Estimation of Executive Compensation by Linear Programming," Management Science 1 (1955), pp. 138-51.

Comiskey, E. E., "Cost Control by Regression Analysis," Accounting Review, April 1966, pp. 235-8.

Ferguson, T. S., "Development of the Decision Model," in On the History of Statistics and Probability, D. B. Owen, editor, 1976, Marcel Dekker Inc., New York, pp. 335-46.

Gauss, K. F., "Theory of the Combination of Observations Which Leads to the Smallest Errors," Gauss Werke, 4, 1821, pp. 1-93.

Ignizio, J. P., "The Development of Cost Estimating Relationships Via Goal Programming," Engineering Economist, Vol. 24, No. 1, 1978, pp. 37-47.

Ijiri, Y., Management Goals and Accounting for Control, North Holland, 1965.

Johnston, J., Statistical Cost Analysis, McGraw Hill, New York, 1960.

Laplace, P. S., Theorie Analytique des Probabilities (Analytic Probabilit Theory), 1812, Courcier, Paris.

Lee, S. M., Goal Programming for Decision Analysis, Auerbach Publishing, 1972.

McClenon, P. R., "Cost Finding Through Multiple Correlation Analysis," Accounting Review, July 1963, pp. 540-7.

Peterson, C. and A. Miller, "Mode, Median, and Mean as Optimal Strategies," Journal of Experimental Psychology, October 1964.

Plackett, R. L., "The Discovery of the Methods of Least Squares," Biometrika, 1972, 59, pp. 239-51.

Sharp, W. F., "Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios," Management Science, Vol. 18, No. 2, 1971, pp. B-1 - B-12.

Wagner, H. M., "Linear Programming Techniques to Regression Analysis," J. of Am. Stat. Assoc., Vol. 54 (1959), pp. 208-12.

Wilson, H. G., "Least Squares Versus Minimum Absolute Deviations Estimation in Linear Models," Decision Sciences, 1978, pp. 322-35.

M/B/162